ANALISIS SENTIMEN DAN PEMODELAN TOPIK OPINI MENGENAI CORETAX MENGGUNAKAN MODEL HYBRID VADER, TF-IDF, BERT DAN BERTOPIC

SKRIPSI

Cindy Febriani Santoso

00000059735

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS TEKNIK DAN INFORMATIKA

UNIVERSITAS MULTIMEDIA NUSANTARA

TANGERANG

2025

ANALISIS SENTIMEN DAN PEMODELAN TOPIK OPINI MENGENAI CORETAX MENGGUNAKAN MODEL HYBRID VADER, TF-IDF, BERT DAN BERTOPIC

SKRIPSI

Diajukan sebagai Salah Satu Syarat untuk Memperoleh

Gelar Sarjana Komputer

Cindy Febriani Santoso

00000059735

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS TEKNIK DAN INFORMATIKA

UNIVERSITAS MULTIMEDIA NUSANTARA

TANGERANG

2025

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Cindy Febriani Santoso

Nomor Induk Mahasiswa : 00000059735

Program Studi : Sistem Informasi

Skripsi dengan judul:

ANALISIS SENTIMEN DAN PEMODELAN TOPIK OPINI MENGENAI CORETAX MENGGUNAKAN MODEL HYBRID VADER, TF-IDF, BERT DAN BERTOPIC

Merupakan hasil karya saya sendiri bukan plagiat dari laporan karya tulis ilmiah yang ditulis oleh orang lain, dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan maupun dalam penulisan laporan karya tulis ilmiah, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk mata kuliah yang telah saya tempuh.

Tangerang, 25 Mei 2025

(Cindy Febriani Santoso)

HALAMAN PERSETUJUAN

Skripsi dengan judul

ANALISIS SENTIMEN DAN PEMODELAN TOPIK OPINI MENGENAI CORETAX MENGGUNAKAN MODEL HYBRID VADER, TF-IDF, BERT DAN BERTOPIC

Oleh

Nama	: Cindy Febriani Santoso
NIM	: 00000059735
Program Studi	: Sistem Informasi
Fakultas	: Teknik dan Informatika

Telah disetujui untuk diajukan pada

Sidang Ujian Skripsi Universitas Multimedia Nusantara

Tangerang, 4 Juni 2025

Pembimbing

Ahmad Faza, S.Kom., M.T.I.

Ketua Program Studi Sistem Informasi

Ririn Ikana Desanti, S.Kom., M.Kom.

HALAMAN PENGESAHAN

Skripsi dengan judul

ANALISIS SENTIMEN DAN PEMODELAN TOPIK OPINI MENGENAI CORETAX MENGGUNAKAN MODEL HYBRID VADER, TF-IDF, BERT DAN BERTOPIC

Oleh

Nama	: Cindy Febriani Santoso
NIM	: 00000059735
Program Studi	: Sistem Informasi
Fakultas	: Teknik dan Informatika

Telah diujikan pada Jumat, 20 Juni 2025
Pukul 15.00 s.d 17.00 dan dinyatakan
LULUS
Dengan susunan penguji sebagai berikut.

Ketua Sidang

01/07 2025

Jansen Wiratama, S.Kom., M.Kom.
0409019301

Penguji

1/7.25

Melissa Indah Fianty, S.Kom., MMSI.
0313019201

Pembimbing

1/7/2025

Ahmad Faza, S.Kom., M.T.I.
0312019501

Ketua Program Studi Sistem Informasi

1/7'25

Ririn Ikana Desanti, S.Kom., M.Kom.
0313058001

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Yang bertanda tangan di bawah ini:

Nama : Cindy Febriani Santoso

NIM : 00000059735

Program Studi : Sistem Informasi

Jenjang : S1

Judul Karya Ilmiah :

ANALISIS SENTIMEN DAN PEMODELAN TOPIK OPINI MENGENAI CORETAX MENGGUNAKAN MODEL HYBRID VADER, TF-IDF, BERT DAN BERTOPIC

Menyatakan dengan sesungguhnya bahwa saya bersedia* **(pilih salah satu):**

☑ Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya ke dalam repositori Knowledge Center sehingga dapat diakses oleh Sivitas Akademika UMN/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial.

☐ Saya tidak bersedia mempublikasikan hasil karya ilmiah ini ke dalam repositori Knowledge Center, dikarenakan: dalam proses pengajuan publikasi ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*) **.

☐ Lainnya, pilih salah satu:

☐ Hanya dapat diakses secara internal Universitas Multimedia Nusantara

☐ Embargo publikasi karya ilmiah dalam kurun waktu 3 tahun.

Tangerang, 25 Mei 2025

(Cindy Febriani Santoso)

KATA PENGANTAR

Puji Syukur penulis sampaikan atas terselesaikannya penelitian dan penulisan skripsi ini dengan judul, "Analisis Sentimen dan Pemodelan Topik Opini mengenai Coretax menggunakan Model Hybrid VADER, TF-IDF, BERT, dan BERTopic." Pembuatan skripsi ini dilakukan untuk memenuhi salah satu syarat mendapat gelar Sarjana Komputer Program Studi Sistem Informasi pada Fakultas Teknik dan Informatika Universitas Multimedia Nusantara. Penulis menyadari bahwa, tanpa bantuan, dukungan, dan bimbingan berbagai pihak sejak memulai masa perkuliahan hingga penyusunan skripsi ini, sangatlah sulit bagi penulis untuk menyelesaikan skripsi ini. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Bapak Dr. Andrey Andoko, selaku Rektor Universitas Multimedia Nusantara.
2. Bapak Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Ibu Ririn Ikana Desanti, S.Kom., M.Kom., selaku Ketua Program Studi Sistem Informasi Universitas Multimedia Nusantara.
4. Bapak Jansen Wiratama, S.Kom., M.Kom., selaku Ketua Sidang yang memberikan kritik dan saran membangun selama proses sidang berlangsung, sehingga penulisan skripsi ini menjadi lebih baik.
5. Ibu Melissa Indah Fianty, S.Kom., MMSI., selaku Penguji yang memberikan kritik dan saran membangun selama proses sidang berlangsung, sehingga penulisan skripsi ini menjadi lebih baik.
6. Bapak Ahmad Faza, S.Kom., M.T.I., sebagai Pembimbing yang telah meluangkan waktu untuk memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
7. Kak Desy dan Mas Ryan yang meluangkan waktunya sebagai narasumber untuk melakukan validasi hasil temuan yang didapat pada skripsi ini.
8. Keluarga penulis sebagai sumber kekuatan dan motivasi yang juga memberikan dukungan secara material, mental, dan moral yang menjadi pendorong utama bagi penulis sehingga dapat menyelesaikan skripsi ini.

9. Teman-teman yang hadir dan mendampingi penulis dalam seluruh dinamika selama menjalani masa perkuliahan dari awal hingga terselesaikannya skripsi ini.

Semoga skripsi ini dapat bermanfaat bagi pembaca dalam memahami respons publik terhadap layanan perpajakan yang diluncurkan pemerintah, berkontribusi pada ilmu berkaitan dengan *Natural Language Processing*, serta menjadi referensi untuk penelitian selanjutnya di masa depan. Penulis juga berharap hasil penelitian dalam skripsi ini dapat memberikan wawasan dan dampak positif, terutama dalam meningkatkan literasi terhadap sistem perpajakan di dunia nyata.

Tangerang, 25 Mei 2025

(Cindy Febriani Santoso)

ANALISIS SENTIMEN DAN PEMODELAN TOPIK OPINI MENGENAI CORETAX MENGGUNAKAN MODEL HYBRID VADER, TF-IDF, BERT DAN BERTOPIC

(Cindy Febriani Santoso)

ABSTRAK

Coretax merupakan sistem perpajakan yang diluncurkan oleh pemerintah Indonesia pada bulan Januari 2025. Sistem ini mengintegrasikan seluruh layanan DJP online ke dalam satu platform. Implementasi sistem Coretax diharapkan dapat meningkatkan penerimaan pajak akibat kepatuhan pajak yang rendah. Sejak perilisannya, sistem ini menghasilkan opini dan reaksi publik di media sosial, terutama X. Penelitian ini melakukan analisis sentimen dan pemodelan topik untuk mengetahui opini publik terhadap Coretax.

Metode analisis sentimen menggabungkan fitur VADER, TF-IDF, dan BERT embeddings ke model klasifikasi Logistic Regression. Sementara itu, pemodelan topik diimplementasikan dengan BERTopic untuk mengekstrak topik yang dibicarakan di X. Penelitian ini menerapkan KDD sebagai kerangka kerja penelitian. Proses pembersihan data terdiri dari penyaringan dan pengubahan teks menjadi huruf kecil, penerjemahan ke bahasa Inggris, tokenisasi, lemmatisasi, dan pemeriksaan ejaan. Di sisi lain, transformasi data dilakukan dengan pelabelan data dan SMOTE untuk menyeimbangkan kelas data. Tahap-tahap sebelumnya menghasilkan 17.100 data valid untuk analisis lebih lanjut.

Penelitian ini menunjukkan bahwa model *hybrid* dapat mengungkap dominasi sentimen negatif terhadap Coretax di Indonesia pada data aktual. Metode penanganan kelas tidak seimbang SMOTE memiliki performa yang tinggi dengan akurasi, presisi, *recall*, dan *f1-score* 94%. BERTopic dapat mengidentifikasi apresiasi fitur yang didiskusikan secara positif dan keluhan teknis yang dihadapi penguna pada sentimen negatif. Validasi hasil mengkonfirmasi relevansi hasil temuan dan analisis kebijakan lebih lanjut menunjukkan inkonsistensi dengan Permenkomdigi Nomor 6 Tahun 2025. Sistem Coretax yang tidak optimal berkontribusi pada penurunan penerimaan pajak sehingga dibutuhkan perbaikan terhadap keluhan pengguna.

Kata kunci: Coretax, sentimen, opini, topik, X

SENTIMENT ANALYSIS AND TOPIC MODELING OF PUBLIC OPINION ON CORETAX USING VADER, TF-IDF, BERT HYBRID MODEL, AND BERTOPIC

(Cindy Febriani Santoso)

ABSTRACT (English)

Newly launched Indonesian government online tax system, Coretax, have just launched in January 2025. It integrates all services in DJP online into a single platform. Coretax is expected to increase tax revenue due to low tax compliance. Since the release date, this new systems led to public opinions and reactions in social media, especially X. This research conducts sentiment analysis and topic modeling to find out public opinion on Coretax implementation.

Sentiment analysis method integrates VADER, TF-IDF, and BERT embeddings features into Logistic Regression model. As for topic modeling implemented using BERTopic to extract what is being discussed in X. This study employs KDD framework as flow of research. Data cleaning include text filtering and case folding, English translation, tokenization, lemmatization, and spellchecking. On the other hand, data transformation is done with data labeling and SMOTE class balancing. These steps resulted in 17.100 tweets for further analysis.

This research shows that proposed hybrid model can reveal the dominance of negative sentiment towards Coretax in Indonesia on actual data. While using SMOTE, the model performs highly with 94% accuracy, precision, recall, and f1-score. BERTopic is able to identify discussion of appreciation positively and technical complaints by users from negative sentiment. Expert validation confirms the relevance of findings. Further policy analysis shows that Coretax does not comply with Permenkomdigi Number 6 of 2025. Inefficiency caused by Coretax has led to decline in tax revenue, making it necessary to address user complaints.

***Kata kunci**: Coretax, opinion, sentiment, topic, X*

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR RUMUS

DAFTAR LAMPIRAN